图书情报工作
第 62 卷 第 7 期 2018 年 4 月

我国图书情报学科在国际上的研究主题特征分析*

石燕青 孙建军
南京大学信息管理学院 南京 210023

摘要：[**目的/意义**]对我国图书情报学在国际权威期刊上发表的文章主题特点进行研究。[**方法/过程**]对文献关键词进行整体分析,根据发文量的变化情况分为三个阶段,梳理出总体研究热点。进一步对三个阶段中发文数量最多、上涨速度最快的阶段进行细粒度划分,基于关键词共现方法分析研究主题的时序变化情况。[**结果/结论**]我国学者在国际上的研究主题集中、多样化程度低,近十年来科学计量学、社会网络主题发展迅速,知识组织和管理主题的研究热度长久不衰。

关键词：共词分析 科学计量学 研究热点 可视化分析
分类号：G250
DOI:10.13266/j.issn.0252-3116.2018.07.008

随着信息泛在化趋势和数据科学的发展,图书情报学的研究不再集中于单一的文献信息,出现了数据多源化、内容多样化、方法跨学科融合化的研究趋势。1992 年召开的全国科技情报工作会议将“科技情报”改为“科技信息”后,更多信息科学群融入图书情报学领域,研究范式也从传统的图书馆信息科学(Library Information Science)独立体向与信息管理科学(Information Management Science)范式的融合演进。科技全球化的环境下,信息科学的国际化趋势带动了我国图书情报学科的国际化发展趋势。邱均平等[1]曾对我国 1916-2011 年的图书情报学国际发文量进行了分析,发现发文量增长迅速,且 70% 以上的论文都集中在 2007-2011 这 5 年中。随着发文数量的增长,我国图书情报学文献的国际影响力也有所提升,有研究发现中国知识转移对象主要是发达国家,且研究主题分布较集中,主要分布在图书情报理论、情报技术应用、学科评价、商业竞争情报 4 个方向[2]。总体来看,我国学者在国际权威期刊上的发文量逐年上升,尤其是国际合作已经成为许多学者发表国际论文的重要模式[3]。为了揭示近年来我国图书情报学领域的国际化发展和研究态势,笔者尝试通过共词分析的方法,对我国图书情报学科在国际上的研究特点进行分析,探索在科技大数据环境下,国内学者的研究模式、研究主题范围、研究前沿等问题,对比更大范围内的国际研究,帮助国内学者进一步明确核心问题,攻克难题。

共词分析法是信息科学中的三大定量方法之一,常常被用于内容分析[4]。2006 年起国内外对共词分析的研究开始增加,包括对学科热点、学科发展过程、研究水平、学科间关系等的研究[5]。关键词是作者对于文章内容的高度凝练概括,某个学科领域在较长一段时间内,大量学术成果的关键词集合能够揭示这些成果的总体内容特征、研究结构和脉络[6],也成为共词分析最常用的词源[7]。不同关键词同时出现在一篇文章中,能说明研究主题间的关系,大量的关键词共现构成一个复杂的网络,共现程度可表示研究主题的热度。因此,笔者选取关键词作为分析研究主题结构的载体,采用关键词共现的分析方法,对我国图书情报学科在国际上的研究特点进行研究。

1 数据来源

笔者以 Web of Science 核心合集为数据来源,采集 1987-2016 年所有包含中国作者参与的图书情报学领域的文献数据,即只要作者地址中出现“中国”的文献都作为命中结果,因此检索式为:SU = Information Science & Library Science AND CU = China,时间跨度:从 1987 至 2016 年,文献类型:Article,共得到 2 817 条数

* 本文系南京大学研究生跨学科科研创新项目“基于本体的社会网络信息语义描述及分析”(项目编号:2016CW07)研究成果之一。
作者简介:石燕青(ORCID:0000-0002-1091-7926),博士研究生,E-mail:yanqings4869@163.com;孙建军(ORCID:0000-0003-2144-8253),院长,教授,博士,博士生导师。
收稿日期:2017-10-12 修回日期:2017-12-17 本文起止页码:66-76 本文责任编辑:王善军

据。Web of Science 中的文献题录中提供了两种关键词:作者关键词和增补关键词,前者是文章作者给出的关键词,后者由 Web of Science 给出。由于增补关键词由标引人员给出,会存在"标引者效应"问题,针对这一问题曾有多位学者进行过研究讨论[8-11]。笔者选取能够有效代表作者观点的"作者关键词"进行分析。对检索到的文献数据进行整理,删除没有关键词的数据,共得到 2 289 篇文献,作为本文的研究对象。

由于作者给出的关键词是非受控词,形式比较主观,对于同一个概念会有不同的表达方式,需先对关键词进行规范化,否则会影响关键词词频,对后续研究起到误导作用[12]。因此,将异形同义词(不同词性、单复数、缩写、有无连字符等)合并,如"co-authorship""co-authors""co-authorship analysis";将近义词合并,如"internet""www"等。此外,删除一些如"China""Hong Kong"等无分析意义的关键词。

2　数据总体描述

首先,将整理好的 2 289 篇文献作整体的描述性统计,对文献量随时间的变化趋势进行可视化分析,得到如图 1 的结果。

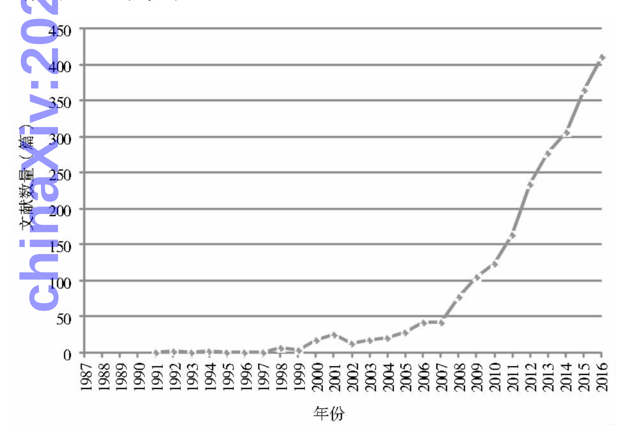


图 1　有关键字的文章总量变化情况

从图 1 中可以看出,剔除无关键词的文献数据后,文献量从 1991 年开始出现。2000 年开始,文章数据超过 10 篇,且上升趋势开始加快。2008 年文章数量超过 50 篇,并以更快的速度增长。根据曲线显示的各年发文量的变化速度,可以将其分为三个阶段:初始阶段(1991－1999)、起步阶段(2000－2007)、发展阶段(2008－2016)。从总体态势来看,这三个阶段的总发文量分别为 17、206、2 066(90.26%),各阶段的发文量较前一阶段均呈指数增长,三个阶段的关键词数量分别为 71、813、6 425(91.27%),这样的增长速度表明我国在国际权威期刊上的影响力不断扩大。

从关键词的角度,对三个阶段分别进行分析。初始阶段,仅有两个关键词"strategy"和"information system"的出现次数为 2,其余关键词的出现次数全部为 1。

起步阶段,在所有的关键词中,出现次数超过 5 次的关键词分别为"internet""e-commerce""knowledge management""information retrieval""digital library""collaboration""multimedia",共现次数超过 1 次的分别为"digital library"和"internet","collaboration"和"knowledge management",与前一阶段相比,起步阶段的研究领域有所拓宽,研究主题更加丰富。可以看出,随着互联网技术的发展,该阶段关于网络、多媒体、数字化、移动终端等主题的研究增加。同时,有关技术、算法的应用类研究增加,定量研究相比上一阶段有所加强。

到了发展阶段,文献量和关键词数量都迅速增长,文献量在三个阶段的占比为 90.3%,关键词数量占比 91.3%。关键词中出现超过 50 次的分别为"bibliometric""social network""citation""social media",值得注意的是,相比前一阶段,该阶段最明显的变化是文献计量学和社会网络的关键词增加,且词频都较高,如"scientometric""h index""network analysis""social network site"等,尤其是文献计量学在前两个阶段基本没有出现,在这一阶段跃居第一梯队,可以看出科学计量和社会网络在当前研究主题中的重要性。这一阶段是文献量增长速度最快的时期,同时也是科技政策深化发展的重要时期。2006 年,我国提出"十一五"期间国际科技合作的战略转变,将科技合作提升至国家战略角度,随着"十二五"专项规划的深化推进和一系列针对国际科技合作的专项政策,国内和国外图书情报学科交流程度加深,我国的国际学术影响力也大大提升,从这一阶段文献总量以及新生的研究节点数量增长情况来看,都具有一定的代表性。对这一阶段国内学者关注的主题进行着重研究,有利于整体把握我国学者在国际范围内关注的学术热点与前沿。

3　科研主题结构分析

从 2008 年起,我国图书情报学界的学者在国际权威期刊上的发文量已超过 50 篇,2009 年已超过 100 篇,并以越来越快的速度增长,根据前文的阶段划分,发展阶段的总文献量和总关键词数量都超过了总数的 90%,因此有必要针对这一阶段的研究现状和主题结构进行进一步的分析。笔者以论文的关键词为研究对

象,通过考察关键词的共现情况,试图对2008到2016年的研究主题结构进行梳理。

3.1 确定分析对象

发展阶段的文献总量为2066篇,关键词总数是6 425个,首先需要确定出高频关键词以便进一步分析。确定高频关键词的方法大概分为几类:①综合考虑关键词数量和词频,根据经验选择高频词[13-14],根据具体情况,词频阈值小至2,大至40的情况都有[15-16];②根据词频变化趋势来决定高频词低频词界限,如马费成等[17]将累积频次在62%以上的作为高频词,王红[18]将词频变化曲线中每个阶段词频大于22的作为高频词;③根据1973年在齐普夫第二定律基础上提出的Donohue高频词低频词界分公式[19],计算高频词阈值。为了避免人工选择带来的主观性,依据Donohue高频词低频词界分公式,得到95个高频关键词,以此作为进一步分析的对象。前50个关键词如表1所示:

表1 高频关键词(部分)

序号	关键词	频次	序号	关键词	频次
1	bibliometric	110	26	data mining	17
2	social network	71	27	virtual community	17
3	citation	64	28	social network site	17
4	social media	51	29	information service	16
5	trust	44	30	patent	16
6	knowledge management	44	31	chinese culture	16
7	digital library	40	32	digital divide	16
8	knowledge sharing	38	33	co-word analysis	15
9	e-commerce	35	34	open access	14
10	case study	30	35	impact factors	14
11	academic library	30	36	information technology	14
12	scientometric	30	37	collaboration network	14
13	internet	28	38	motivation	13
14	h-index	27	39	co-author	13
15	research trend	26	40	evaluation	13
16	gis	25	41	customer satisfaction	13
17	collaboration	24	42	knowledge transfer	13
18	text mining	22	43	satisfaction	13
19	cellular automata	21	44	library	12
20	innovate	21	45	decision support system	12
21	information retrieval	20	46	culture	12
22	cluster	20	47	humanities and social science	12
23	e-government	20	48	structural equation modeling	12
24	network analysis	18	49	classification	12
25	internet of science	18	50	technology acceptance model	12

高频关键词表示了学者们相对集中的研究领域,在一定程度上反映了某一时段该学科的研究热点。不同关键词在同一篇文章中的共现,能够揭示主题间的关系,从而反映研究结果和走向,采取共词分析的方法,对上文得到的95个关键词在所有文章中的共现次数进行统计,得到一个95×95的矩阵,如表2所示(仅显示部分)。

关键词共现的次数受关键词自身频次的影响很大,为了消除这种影响,使用Ochiai系数将高频关键词共现矩阵标转化,转换为相似矩阵。Ochiai系数的计算方式是:XY系数=X、Y同时出现的次数/(X出现的次数×Y出现的次数)$^{1/2}$,标准化后的矩阵消除了因关键词自身频次差异带来的影响,真正表示了关键词间的相关性,数值越大,表明两个关键词的关系越密切。

由表2可以看出,共词矩阵比较稀疏,矩阵中0值较多,为了避免0值给后续分析带来的误差,将上述Ochiai系数相似矩阵转化为相异矩阵,具体方法是用“1”减去相似矩阵,得到的是关键词相异矩阵,如表3所示。相异矩阵表示各关键词的不相似性,数值越大,表明两个关键词的相似性越弱,距离越远。

石燕青，孙建军. 我国图书情报学科在国际上的研究主题特征分析[J]. 图书情报工作,2018,62(7):66－76.

表 2　高频关键词共现矩阵(部分)

	bibliometric	social network	citation	knowledge management	knowledge sharing	case study	scientometric	h-index	research trend	gis	collaboration
bibliometric	110	4	13	1	0	2	5	3	19	1	3
social network	4	71	0	0	0	0	0	0	0	0	0
citation	13	0	64	1	0	0	2	4	0	0	0
knowledge management	1	0	1	44	9	0	1	0	0	0	0
knowledge sharing	0	0	0	9	38	0	0	0	0	0	0
case study	2	0	0	0	0	30	2	0	2	0	0
scientometric	5	0	2	1	0	2	30	0	0	0	0
h-index	3	0	4	0	0	0	0	27	0	0	0
research trend	19	0	0	0	0	2	0	0	26	0	0
gis	1	0	0	0	0	0	0	0	0	25	0
collaboration	3	0	0	0	0	0	0	0	0	0	24

表 3　高频关键词相异矩阵(部分)

	bibliometric	social network	citation	knowledge management	knowledge sharing	case study	scientometric	h-index	gis	collaboration
bibliometric	0	0.954 7	0.845 1	0.985 6	1	0.965 2	0.913	0.945	0.980 9	0.941 6
social network	0.954 7	0	1	1	1	1	1	1	1	1
citation	0.845 1	1	0	0.981 2	1	1	0.954 4	0.903 8	1	1
knowledge management	0.985 6	1	0.981 2	0	0.779 9	1	0.972 5	1	1	1
knowledge sharing	1	1	1	0.779 9	0	1	1	1	1	1
case study	0.965 2	1	1	1	1	0	0.933 3	1	1	1
scientometric	0.913	1	0.954 4	0.972 5	1	0.933 3	0	1	1	1
h-index	0.945	1	0.903 8	1	1	1	1	0	1	1
gis	0.980 9	1	1	1	1	1	1	1	0	1
collaboration	0.941 6	1	1	1	1	1	1	1	1	0

3.2　多维尺度分析

多元统计分析方法是共词分析的核心内容，常用的多元统计方法有三种：因子分析、聚类分析、多维尺度分析。多维尺度分析通过模型拟合，将多个变量映射到二维或三维空间，采用可视化方法来展示数据间的关系，通过多维尺度分析，可以比较直观地观察到某学科内的研究热点聚集情况、研究主题结构等信息。

一般来讲，相异矩阵最适合用多维尺度分析方法。因此，首先采取多维尺度分析法对上文得到的相异矩阵进行分析，观察研究主题的整体分布状况。由于相异矩阵是由 Ochiai 相似矩阵转换而来，矩阵中的数据表示不同关键词间的共现程度，能代表各关键词的距离，数字越大，代表共现次数越少、距离越远。因此，采用 SPSS 软件中的多维尺度分析(ALSCAL)，设置"数据为距离数据"，选择正对称形状，用序数数据作为测度水平，对相异矩阵进行二维分析，得到的结果见图 2。

由模型的拟合统计量可以看到，RSQ 是决定系数，代表变异中能够被模型解释的程度，数值为0.962 94，说明模型拟合效果较好。坐标图显示出本文所研究的论文主题概貌，从各个关键词构成的距离分布上来看，并未很明显地分成几大类别，这表明我国学者在国际期刊上所发文章的主题比较集中，没有形成显著不同的主题派系，研究不够多元化。各个关键词在整体上虽然没有形成明显的主题模块划分，但依然可以根据分布距离的远近程度，大致分为三块内容：

3.2.1　科学计量主题　第一象限的关键词主要包含科学计量，如"bibliometric""citation""h-index""scientometric""co-word analysis""impact factors"等，相较于其他主题，这一主题的关键词频率也较高，这表明我国图书情报学科的学者在国际权威期刊上发表的文章中，偏向数据分析的文章更多，科学计量主题占了较大比重；除了科学计量，科学合作类关键词在这一区域出现的也比较多，如"co-author""collaboration network""international collaboration""interdisciplinary"等，随着网络技术的发展，在 E-science 环境下，为了得到更多

图书情报工作
第 62 卷 第 7 期　2018 年 4 月

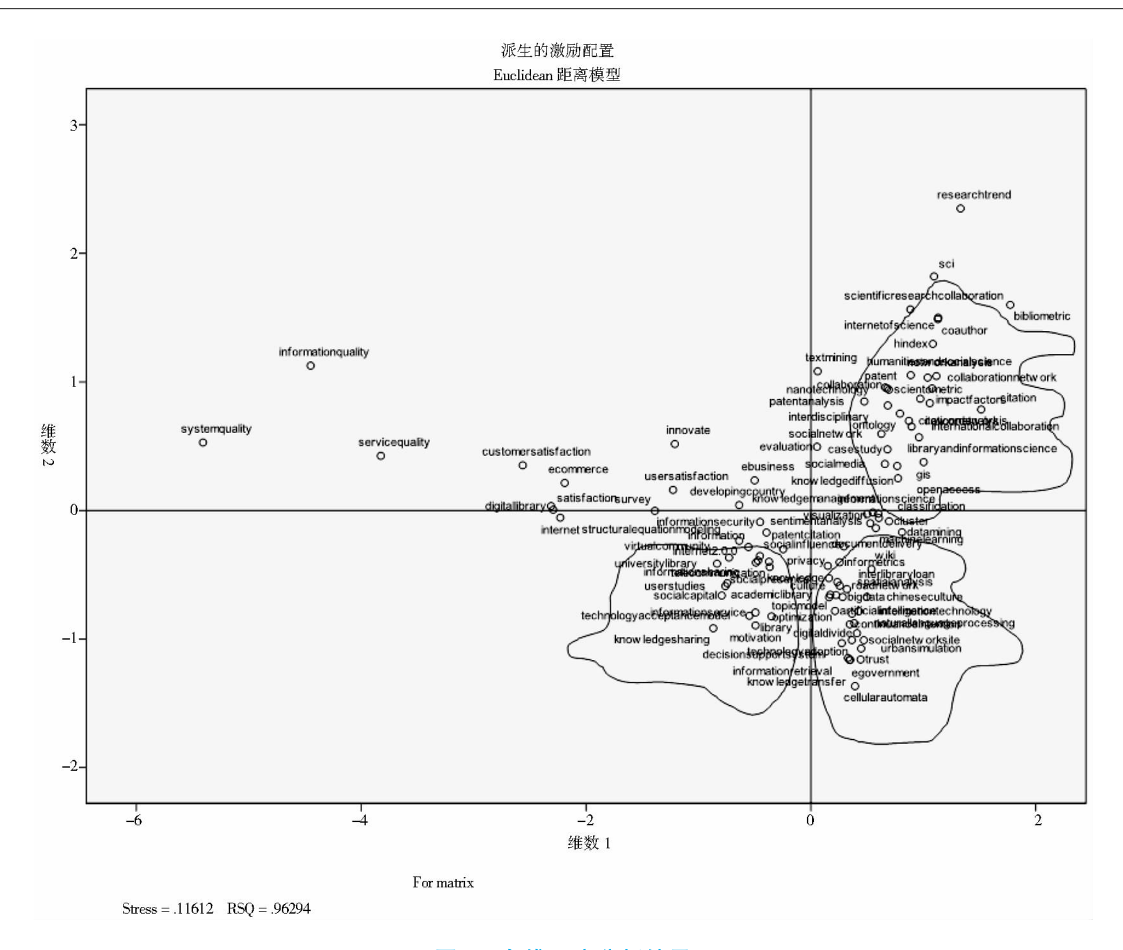


图 2　多维尺度分析结果

的信息资源，降低科研成本，更有效地创造科学价值，科学研究逐渐从个体研究转向合作研究，在这种模式下，针对合作行为的研究也逐渐增加，这方面的研究从定性到定量都有，从图 2 中这一主题与科学计量主题的关系来看，我国学者在这方面的文章更偏向定量研究。

3.2.2　数据分析主题　第二象限的关键词包括：“social network site”“road network”“spatial analysis”等网络分析、空间分析及可视化内容，“big data”“data mining”“cluster”“classification”等数据分析方法，此外还包括“machine learning”“artificial intelligence”“natural language processing”等技术导向的关键词。其中，数据分析方法的关键词分布于第一和第二象限交界处，说明科学计量、针对学术关系的研究、网络结果分析等领域都用到了聚类、可视化等数据分析法，这些方法仍是目前图书情报学科常用的数据处理方式。

3.2.3　知识管理主题　图 2 中的第三象限中，如“information security”“digital divide”“information service”“information sharing”等，侧重于讨论信息服务与管理，此外还包含“library”“university library”“academic library”等图书馆研究方面的关键词，“users study”“sentiment analysis”“virtual community”等针对用户的研究等。在这部分关键词中，较少出现技术导向和数据分析类关键词。

第四象限的关键词数量最少，分布也最稀疏，其中“digital library”和“satisfaction”位于三、四象限的交界，与上文中“知识管理”主题很接近。

总体来看，依据图 2 的聚类划分结果，虽然各主题模块之间的区分界限并不明显，但相对而言，类别内的关键词距离更近，类别间的关键词距离更远。综合关键词的距离分布及高频关键词表来看，我国图书情报学领域的学者在国际权威期刊上所发表的文章中，文献计量和科学计量是最主要的一个主题，与该主题结合进行的研究中，科研合作关系研究所占比重较大，在科技全球化的背景和逐渐发展的 E-science 环境下，科研合作已经成为一种趋势，运用计量学知识对合作关系进行研究，是一种较直观且方便的方式。

此外，在对线上社交网络以及其他新技术进行探讨的主题中，多结合数据分析方法，在一定程度上可以看出，定量分析是我国学者在国际权威期刊上发文时

偏爱的研究方法。信息服务与质量、资源共享方面的主题多结合用户分析来进行研究，这方面的主题与数据分析和技术类关键词的距离相对较远，多偏向定性研究。

3.3 聚类分析

聚类分析是一种重要的多元统计分析法，聚类是根据变量的相似性，将关系密切的变量聚集为一类，聚类分析本质上是一种统计描述方法，聚类结果没有统计检验的方法来验证，知识根据变量自身的信息，使得类别内的变量差异尽可能小，类别间的变量差异尽可能大，常用的方法是层次聚类法。

经过多维尺度分析，能够大致将研究主题划分为三个部分，为了对研究主题进行更详细的类团划分，对关键词进行系统聚类分析。经过比较，选择组间联接和欧式平方距离作为聚类方法和度量标准，获得了较好的聚类结果，在类团的命名上，采用钟金伟提出的粘合力指标[20]，粘合力测量了类团中每个元素与其他元素之间的共现次数，揭示了元素在类团中的重要程度，类团划分结果如表 4 所示：

表 4 类团划分结构

序号	类团名称	包含元素个数	类团元素及其粘合力
1	digital library	13	service quality 1.417, informationquality 1.500, system quality 1.417, evaluation 0.500, library 0.333, document delivery 0.667, interlibrary loan 0.500, digital library 2.083, academic library 1.083, information, service 1.333, technology acceptance model 0.167, customer satisfaction 0.917, user studies 0.583
2	collaboration	11	collaboration 1.100, cluster 0.300, network analysis 1.000, patent 1.000, collaboration network 0.700, co-author 0.600, humanities and social science 0.200, international collaboration 0.600, scientific research collaboration 0.600, patent citation 0.100, nanotechnology 0.600
3	bibliometric	13	bibliometric 5.750, scientometric 2.083, research trend 2.667, internet of science 1.750, sci 1.000, citation 2.583, social network analysis 1.500, h-index 1.500, co-word analysis 1.000, impact factors 0.917, big data 0.333, knowledge diffusion 0.833, library and information science 0.417
4	information retrieval	10	knowledge management 1.000, internet 1.000, information retrieval 1.333, text mining 1.000, classification 0.556, natural language processing 1.222, search engine 0.444, artificial intelligence 0.556, machine learning 0.667, patent analysis 0.222
5	business intelligence	5	data mining 1.450, decision support system 0.500, business intelligence 1.500, spatial analysis 0.250, information management 0.500
6	information science & visualization	3	citation network 1.000, visualization 1.500, information science 1.500
7	social media	5	social media 2.750, social network site 2.250, continuance intention 1.250, social influence 0.750, sentiment analysis 0.500
8	Trust & knowledge sharing	9	trust 2.750, social network 1.375, knowledge sharing 2.125, e-commerce 1.375, social capital 1.250, virtual community 1.625, motivation 1.125, structural equation modeling 0.625, satisfaction 1.000
9	electronic commerce	5	electronic commerce 1.500 technology adoption, 0.750, e-business 0.750, online shopping 0.750, intrinsic motivation 0.250
10	information technology	21	case study 0.150, gis 0.150, cellular automata 0.200, innovate 0.200, e-government 0.150, information technology 0.250, digital divide 0.150, Chinese culture 0.100, culture 0.200, open access 0.050, telecommunication 0.200, knowledge transfer 0.150, social presence 0.100, ontology 0.050, university library 0.100, information 0.100, wiki 0.050, optimization 0.000, road network 0.000, information system 0.050, informetrics 0.100

通过聚类分析，最终得到 10 个类团。类团 1 中共有 10 个关键词，包括“library”“digital library”“academic library”等各类图书馆，以及“service quality”“system quality”“customer satisfaction”“user studies”等关于用户服务的关键词。图书馆是图书情报学的一个重要研究对象，随着互联网技术的发展，图书馆的数字化已成为热点研究，数字化进程不仅体现了图书馆的资源集聚能力，而且展现了知识价值挖掘与塑造的过程，在这一基础之上实现的知识发现，也重新定义图书馆的用户服务工作，对于图书馆服务方面的研究，进一步促使了图书馆释放出更多的服务功能。类团 2 包含的关键词主要是网络分析方面，如合作网络、引用网络、专利网络等，在科技全球化的趋势下，合作已成为开展科研活动的重要方式，探究我国与不同国家和地区的国际科研合作模式探究[21－23]，一直是图书情报学的一个研究热点。类团 3 是图书情报领域最重要的研究主题之一，关键词“bibliometric”的粘合力及其与其他关键词的共现次数表明了它在该类团中的重要程度，科学计量作为图书情报学科的重要研究方法，能够应用于气象学、生物学[24]等各个领域的研究中。此外该类团中还包含了“citation”“h-index”“impact factors”等计量指标，从多维视角下的分析共引网络[25]、基于引文的知识扩散研究[26]以及其他

基于计量指标的文献研究都受到了广泛的关注。类团 4 包含“information retrieval”“classification”“artificial intelligence”等数据处理技术;类团 6 中的可视化技术涉及到图像处理、计算机图形学等多个领域,是科学研究中最重要的研究方法之一。类团 7 是社交网络[27]主题,基于社交网络的社会关系、用户行为研究逐渐成为近年来图情领域的研究热点。类团 8 包含虚拟社区、知识共享、社会资本等主题,类团中的元素粘合力均比较强,表明这些研究联系紧密。类团 9 包含的电子商务在我国一直保持着较快的发展态势,且潜力巨大,如何运用网络策略更有效地建立用户信任[28]、在我国的现有环境下基于关系的市场如何从战略管理角度调整电子商务模式[29],得到了学者的广泛关注。

为了进一步分析各个类团揭示的研究主题的发展状况,分别计算每个类团在网络中的密度和中心度,以各个类团内部的实际关系数与理论最大关系数的比值作为类团的密度,各类团内部的元素与类团外部元素的共现次数作为类团的中心度。得到各类团的密度和中心度如表 5 所示:

表 5　各类团的密度和中心度

序号	类团	密度	中心度
1	digital library	0.961 5	86
2	collaboration	0.618 2	79
3	bibliometric	1.717 9	124
4	information retrieval	0.8	105
5	business intelligence	0.9	27
6	information science &visualization	1.333 3	13
7	social media	1.5	40
8	trust	1.472 2	116
9	electronic commerce	0.8	32
10	information technology	0.12	96
	平均值	1.022 3	71.8

将各类团的平均密度和平均中心度作为原点,绘制战略坐标图,以分析各个类团在整个网络中的重要程度,如图 3 所示:

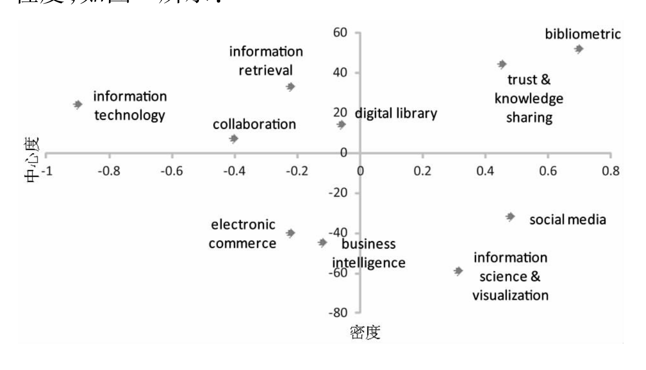


图 3　各类团的战略坐标

从战略坐标图中可以看出,“trust&knowledge sharing”“bibliometric”位于第一象限,密度和中心度都比较高,说明这两个类团内部联系紧密,相关主题的研究发展成熟,同时与外部元素关系也比较密切,尤其是科学计量类团,在整个网络中处于核心地位;有四个类团位于第二象限,分别是“digital library”“information retrieval”“collaboration”“information technology”,其中“digital library”距离原点很近,类团与外部元素的联系程度和类团内部联系程度都处于中间水平,“information retrieval”类团中包含一些数据处理方法,内部联系紧密,“information technology”密度较高,该主题研究已发展得相对成熟,但与外部联系较少,如长期缺乏外部动力,对该主题的进一步发展不利;“electronic technology”“business intelligence”位于第三象限,内部元素联系松散,且与其他研究主题相关度不高,是网络中的边缘类团;“information science & visualization”和“social media”位于第四象限,中心度较高,密度较低,与其他研究主题联系紧密,在网络中处于比较重要的位置,但是类团内部联系较为松散,在未来的发展中有可能被分散到其他的类团主题中。

3.4　研究主题时序变化研究

为了进一步对该阶段的研究主题进行细化研究,以三年为一个时期,将“发展阶段”划分为三个时期,分别进行主题热点演化的可视化分析。

2008 – 2010 年、2011 – 2013 年、2014 – 2016 年为第一、二、三阶段,分别得到的文献数量为 612、677、1 083,关键字数量为 1 173、2 525、3 772。由于各阶段的低频词数量差距过大,因此这里选择词频 > 5 的词作为高频关键词,删除无意义的词,分别得到三个阶段的高频词数量为 48、38、62,分别对其进行共现次数统计,用 pajek 对共现矩阵进行可视化。

图 4 展示了发展阶段第一时期的关键词共现情况,点代表关键词,边代表两个关键词的共现关系,边的粗细表示权重,代表共现次数的多少,点的边数越多且边的权重越大表明该点在网络中越重要。从图 4 中可以看出,依据重要程度排在前几位的分别是“digital library”“information service”“customer satisfaction”“electronic media”“academic library”“computer application”等,可见学者关注的主要是信息数字化、信息服务能力、用户服务等主题,该阶段出现的技术类关键词较少,“scientometric”“data mining”等数据分析方法的关键词也处于边缘位置,说明在该阶段我国学者在国际权威期刊发表的研究多偏向质性分析。

图 4　2008－2010 年高频关键词共现网络

从图 5 的共现网络中可以看到，这一阶段的研究主题与上一阶段大相径庭。边数和边权重较大的关键词分别是“bibliometric”“social network”“knowledge sharing”“scientometric”“citation”等，文献计量学类的关键词在网络中占据很重要的位置，另一个关键词“social network”的 13 条边中，有 5 条与文献计量类关键词相连，与“network analysis”共同出现的也都是计量或数据处理的关键词，可见这一时期我国学者的关注点逐渐专向计量学研究，“data mining”等数据分析类关键词也更多地出现在研究中，定量分析从这一阶段开始逐渐增多。同时，社交网络中的知识管理和共享仍是研究的另一个重点，与上一阶段类似，这类研究仍旧大多属于定性研究，与定量方法结合不密切。

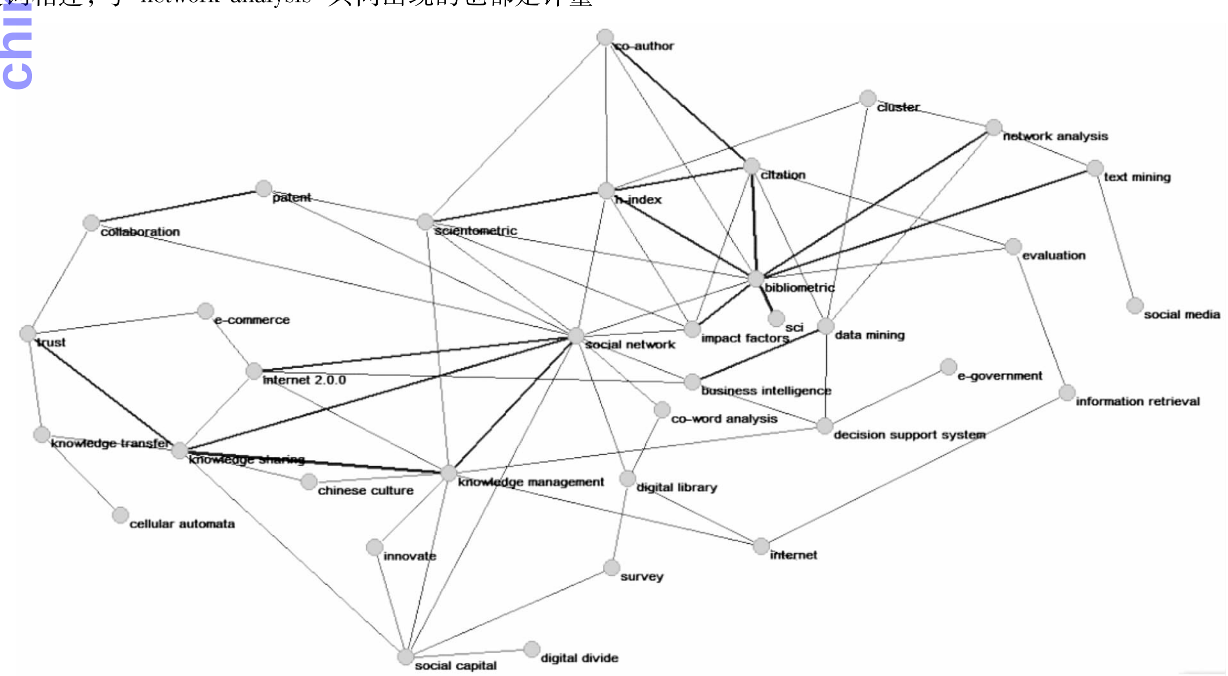


图 5　2011－2013 年高频关键词共现网络

2014－2016 年是发展阶段的第三时期，这一时期的关键词数量最多，依据上文得到的高频词画出的网络图非常复杂，为了简化网络结构，更清晰地看到词间的关系，在原来的基础上，以边数 > 3 且边的权重 ≥3 为条件删除了一部分节点，重新得到共现网络，如图 6 所示。从图 6 中可以看到，该时期的关键词关系是发展阶段中最复杂的，研究主题大致可以分为三类：文献计量、网络分析、信息和知识管理。"bibliometric"是网络中最重要的节点，与之关系最密切的关键词是"research trend"，可见使用文献计量学来分析研究趋势已经成为一种重要方法，其他与之共现的词大多是计量类关键词和网络分析类关键词。在网络分析中，最主要的研究对象是社交网络和合作网络，在这一时期对于网络分析更多地使用到了计量学方法，在前两个阶段定性研究的基础上，增加了定量研究方法。

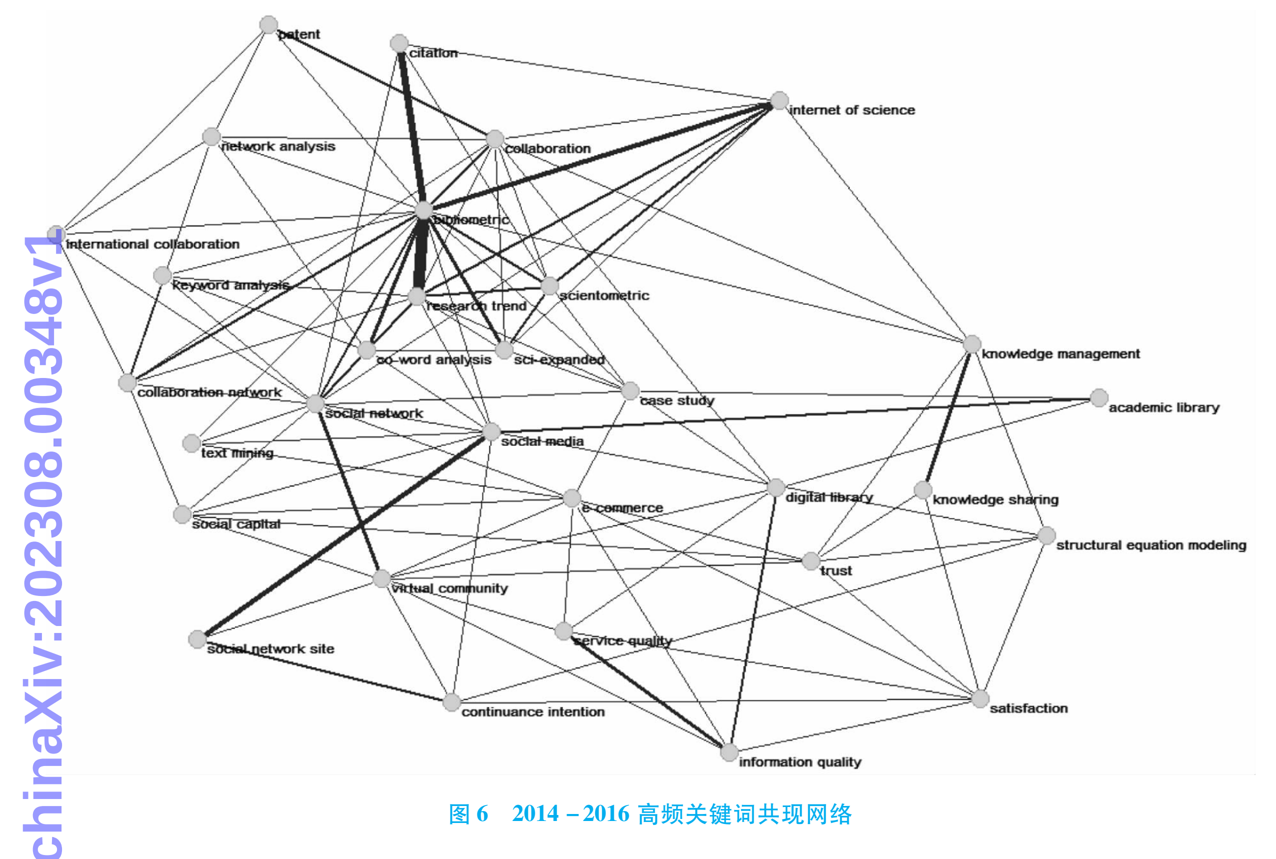


图 6　2014－2016 高频关键词共现网络

纵观发展阶段三个时期的关键词共现网络图，有如下发现：

（1）数据驱动研究逐渐成为国际学术界的主流方法。在整个发展阶段中，出现频次最高的科学计量类关键词是从该阶段的中后期开始出现并逐渐发展起来的。从第一时期非常边缘的位置，到第二时期占据网络中心位置，再到第三阶段重要程度的巩固，计量学迅速成为整个发展阶段出现词频和共现词频最高的关键词。这反映出我国学者在国际上的研究，逐渐从定性方法到定量方法的过渡，科学和文献计量学作为图书情报学科特有的方法，能够对文献和科学数据进行充分的利用，进一步挖掘文献信息。但是结合战略图分析，我国学者采用的定量研究过多集中于计量学方法，对于文本挖掘、自然语言处理、机器学习、人工智能等已经发展得较为成熟的数据分析方法，应用不够广泛。

（2）社交网络研究发展迅速。发展阶段的另一大研究主题"社交网络"，由该阶段的初期开始出现，到中后期开始逐渐变成研究重点。第一时期"social network"在网络中的重要程度属于中间梯队，相关研究方法简单，共现的词只有"cluster"，相关联的研究主题大多属于知识管理方面；第二时期研究主题更加丰富，包含合作、专利、商业情报、数字图书馆等网络，研究方法更多样化，计量学方法占主要地位；第三时期已跃居第一梯队，研究主题和方法都有所丰富，前两个时期处于边缘位置的"social media"也成为网络中的重要节点。但是另一方面，我国图情学者在社交网络主题方面的研究，虽然涉及的领域众多，使得社交网络成为研究热点，但并未形成稳定成熟的研究方法体系。

（3）知识的组织和管理是始终贯穿整个发展阶段的研究主题。从三个时期的关键词发展趋势可以看

出,“knowledge management”“knowledge sharing”始终是网络中的重要节点,即使到了中后期,社交网络和计量学异军突起,知识组织和管理仍旧是整个网络中不可忽视的主题。前期主要借助电子媒体、计算机技术,侧重于信息服务和用户体验;中期随着社交网络的发展,知识管理也更多地结合社交网络进行研究;到了后期,知识管理的研究与计量学结合更加紧密。从某种程度上,知识管理主题的发展也体现了图书情报学的研究本质,现代科技的发展为我们提供了更加多样化的方法、更加丰富的研究对象,但图书情报学的研究始终离不开对信息和知识的组织、挖掘和管理。

4 结语

笔者以 Web of Science 核心合集为数据来源,对我国学者研究的主题进行梳理,并针对发文最集中的阶段进行了细化的时序分析。经过分析,发现我国学者在国际权威期刊上发表的文章主题比较集中,没有形成多个较为独立的研究主题板块,研究多元化程度不高,尤其是近年来的研究主要集中在社交网络和计量学主题。

首先,随着信息泛在化和社交媒体的快速发展,泛社交成为趋势,信息量也随之剧增,这些海量数据中孕育着价值潜力巨大的信息,作为信息研究学科,图书情报学者已成为社交网络研究的主要群体。然而,经过分析发现,我国图书情报学者虽然对这一主题的研究数量较多,但研究不够细致深入,呈现出主题缺乏关联、研究方法分散且不成体系的现象,对这一主题还未形成成熟的研究风格。对于社交网络这一重要的信息资源,应更多地结合网络结构、信息传播、用户行为等方向,拓展研究模式,进一步加强理论支撑,进行更深入的验证性研究。

其次,科学和文献计量研究一直是图书情报学界的重要研究方向,文献题录信息是文献内容的浓缩,同时其结构化的特征使得分析更加方便可靠,因此计量成为学者们钟爱的方法。目前计量学不仅用来分析文献,对于社交媒体上的各类标签也同样适用。然而,计量学仅仅根据已有的题录条目进行分析,对于内容的挖掘能力有限,如果要对文献以及其他类型的信息进行更加深入的分析,我国图书情报学者应突破本领域的方法局限,将目光投向机器学习、人工智能、文本发掘等科研方法,通过方法结合、思路创新等方式扩宽本学科的科研路径。

最后,知识管理主题始终是学者们关注的重点,也是图书情报学科的信息增长点。从上文的战略图中可以看出,知识管理和共享主题不仅是一个发展成熟的学科,与各主题间的关联也非常紧密,作为一个以信息为主要研究对象的学科,知识管理也是图书情报学的重要阵地,有效地对信息进行组织、管理、提炼以获得有价值的知识是学者们的共同目标。在这一领域的研究中,我国图情学者应该以本学科理论和知识管理理论为基础,以多样化技术为支撑,吸收跨学科(如计算机、社会学)的研究方法,同时关注新技术和智能化知识管理的趋势,以进一步巩固知识管理主题的研究。

本文仅仅进行了探索性研究,研究结果仍需综合多种方法进一步验证,对于研究主题的趋势发展也需要深入展开讨论。此外,为了更全面地分析研究主题的分布和发展,需要考虑国内外整体学科的研究趋势,以便进行比较分析,得出更加有价值和指导意义的结论,这些仍需在将来的研究中进一步分析。

参考文献:

[1] 邱均平,曾倩. 我国图书情报学国际发文的计量分析[J]. 图书与情报,2013(1):90－97.

[2] 宋凯,李秀霞,赵思喆,等. 我国图书情报学文献的国际知识影响力分析[J]. 情报理论与实践,2017,40(7):38－42,55.

[3] 赵蓉英,魏绪秋. 基于比较静态分析的国际合作与交流特征演化研究——以我国图书情报学为例[J]. 图书馆,2017(8):27－31,49.

[4] RONDAPUP GA, GUERRAS MARTIN LA. Dynamics of the evolution of the strategy concept 1962－2008: a co-word analysis[J]. Strategic management journal, 2012,33(2): 162－188.

[5] 冯璐,冷伏海. 共词分析方法理论进展[J]. 中国图书馆学报,2006,32(2):88－92.

[6] 李文兰,杨祖国. 中国情报学期刊论文关键词词频分析[J]. 情报科学,2005,23(1):68－70,143.

[7] 胡昌平,陈果. 科技论文关键词特征及其对共词分析的影响[J]. 情报学报,2014,33(1): 23－32.

[8] LAW J, WHITTAKER J. Mapping acidification research: a test of the co-word method[J]. Scientometrics, 1992,23(3):417－461.

[9] CALLON M, COURTIAL J P, TURNER W A. From translation to problematic network: an introduction to co-word analysis [J]. Social science information,1983,22(2):191－235.

[10] WHITTAKER J. Creativity and conformity in science: titles, keywords and co-word analysis[J]. Social studies of science,1989,19(3):473－496.

[11] KING J. A review of bibliometric and other science indicators and their role research evolution[J]. Journal of information science, 1987,13(5):261－276.

[12] 邵作运,李秀霞. 共词分析中作者关键词规范化研究——以图书馆个性化信息服务研究为例[J]. 情报科学, 2012(5): 731

－735.
[13] 张勤,马费成. 国外知识管理研究模式——以共词分析为方法[J]. 科学管理学报,2008,10(6):65－75.
[14] 李武,董伟. 国内开放存取的热点研究:基于共词分析的文献计量研究[J]. 中国图书馆学报,2010,36(6):105－115.
[15] DING Y,CHOWDHURY G G,FOO S. Bibliometric cartography of information retrieval research by using co-word analysis[J]. Information processing & management,2001,37(6):817－842.
[16] 邱均平,丁俊达,周春雷. 1999－2008年我国图书馆学研究的实证分析(上)[J]. 中国图书馆学报,2009,35(5):72－79.
[17] 马费成,望俊成,陈金霞,等. 我国数字信息资源研究的热点领域:共词分析透视[J]. 情报理论与实践,2007,30(4):438－443.
[18] 王红. 近十年我国图书情报学科研究热点的共词分析[J]. 情报学报,2011,30(7):765－775.
[19] DONOHUE J C. Understanding scientific literature:a bibliographic approach[M]. Cambridge:TheMITPress,1973.
[20] 钟伟金,李佳. 公司分析法研究(二)——类团[J]. 情报杂志,2008,27(6):141－143.
[21] HE T W. International scientific collaboration of China with the G7 countries[J]. Scientometrics,2009,80(3):571－582.
[22] TANG L,SHAPIRA P. China-US scientific collaboration in nanotechnology: patterns and dynamics[J]. Scientometrics, 2011, 88(1):1－16.
[23] ZHOU P,GLANZEL W. In-depth analysis on China's international cooperation in science[J]. Scientometrics, 2010, 82(3): 597－612.
[24] LI L L,DING G H,FENG N,et al. Globa lstem cell research trend: bibliometric analysis as a tool for mapping of trends from 1991 to 2006[J]. Scientometrics,2009,80(1):39－58.
[25] CHEN C M,IBEKWE-SANJUAN F,HOU J H. The structure and dynamics of cocitation clusters:a multiple-perspective cocitation analysis[J]. Journal of the American Society for Information Science and Technology,2010,61(7):1386－1409.
[26] LIU Y X,ROUSSEAU R. Knowledge diffusion through publications and citations:a case study using ESI-fields as unit of diffusion[J]. Journalof the American Society for Information Science and Technology,2010,61(2):340－351.
[27] LIN H,FAN W G,CHAU P Y K. Determinants of users' continuance of socia lnetworking sites:a self-regulation perspective[J]. Information & management,2014,51(5):595－603.
[28] SIA C L,LIM K H,LEUNG K,et al. Web strategies to promote Internet shopping:is cultural-customization needed? [J]. MIS quarterly,2009,33(3):491－512.
[29] MARTINSONS M G. Relationship-based e-commerce: theory and evidence from China[J]. Information systems journal, 2008, 18(4):331－356.

作者贡献说明:
石燕青:数据采集、处理及分析,撰写论文;
孙建军:提供部分文献资料,指导写作。

Analysis on the Theme Structure and Features of Chinese Library and Information Science in International Studies

Shi Yanqing　Sun Jianjun
School of Information Management, Nanjing University, Nanjing 210023

Abstract: [**Purpose/significance**] This paper analyzes the theme structure and features of Chinese library and information science articles published in international authoritative journals. [**Method/process**] This paper analyzed the literature keywords, it divided the research into three stages according to the changes in the number of annual papers, and combed out the overall research topics. Then, we chose the stage with maximum number of papers and the fastest rising speed of publishing, and broke it into several phases to analyze the sequential change of topics by the method of keyword co-occurrence. [**Result/conclusion**] The research topics of our country scholar are relatively concentrated and less diverse. Scientometrics and social network theme analysis have developed rapidly in the past ten years, and the research themes on knowledge organization and management have been popular.

Keywords: co-words analysis　scientometrics　research hotpot　visual analysis